63 4-2

Department of Statistics
UNIVERSITY OF WISCONSIN
Madison, Wisconsin

Technical Report No. 14
January, 1963

BAYES' THEOREM AND THE USE OF PRIOR KNOWLEDGE IN REGRESSION ANALYSIS

George C. Tiao and Arnold Zellner

This research was supported in part by the United States Navy through the Office of Naval Research, under Contract Nonr-1202(17), Project NR 042 222.

Also issued as Systems Formulation and Methodology Workshop Paper No. 6301 by the Social Systems Research Institute, University of Wisconsin.

Bayes' Theorem And The Use Of Prior Knowledge In Regression Analysis

George C. Tiao and Arnold Zellner

I. INTRODUCTION

The use of Bayes' theorem in statistical inference has recently been reconsidered in the works of Jeffreys (1957, 1961), Savage (1959, 1961, 1962), Raiffa and Schlaifer (1962), Box and Tiao (1962, 1963) and others. Emerging from these works are what we consider to be at least two distinct advantages of the Bayesian approach. First, this approach provides an excellent framework for the systematic and logical assessment of the adequacy of the assumptions which are used in many statistical models. Examples which illustrate this use of the approach may be found in the works of Box and Tiao in which the effects of certain departures from normality are assessed in making inferences about location and scale parameters. Second, given that a model is adequate, the Bayesian approach is one in which prior knowledge about parameters of interest can be combined in a well-defined mathematical way with information obtained from an experiment. Such prior knowledge, which may arise from general theoretical considerations and/or the results of previous or concurrent experiments, is usually an important component of an investigator's quest for understanding. In this paper we illustrate how prior knowledge can be utilized in conjunction with sample information in making inferences about the parameters of the regression model, a model which is used extensively in many areas of research.

The plan of the paper is as follows. In Section 2, we review several Bayesian analyses of the regression model which have appeared in the literature and go on to develop two additional models which we believe

have desirable features not found in other models. Some technical results needed to implement the models in practice are presented in Section 3. Then in Section 4 we apply our methods in the analysis of investment data relating to two large corporations. Finally, in Section 5 we provide a summary.

II. BAYESIAN ANALYSIS OF THE REGRESSION MODEL

2.1 Specification of the Model

We employ the Bayesian approach to make inferences about a regression coefficient vector $\beta' = (\beta^1, \beta^2, \ldots, \beta^p)$. This vector of coefficients appears in the usual regression model as follows:

$$y = X\beta + \epsilon \tag{2.1}$$

where y is a Tx1 vector of observations, X is a Txp matrix of fixed elements with rank p, and ϵ is a Tx1 vector of random disturbances. We assume that the elements of ϵ are normally and independently distributed, each with mean zero and unknown variance σ^2. Under these assumptions our joint likelihood function is:

$$\ell(\beta, \sigma|y) = (1/\sigma\sqrt{2\pi})^T \exp\left\{-(1/2\sigma^2)(y - X\beta)'(y - X\beta)\right\} . \tag{2.2}$$

For simplicity in notation we shall use the symbol $Q(\beta, \eta, A)$ throughout this paper to denote a quadratic form in variables β centered at η and with matrix A, namely

$$Q(\beta, \eta, A) = (\beta - \eta)' A (\beta - \eta).$$

In this notation, the likelihood function can be written:

$$\ell(\beta, \sigma|y) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^T \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + Q(\beta, \hat{\beta}, Z)\right]\right\} \tag{2.3}$$

where $Z = X'X$, $\hat{\beta} = Z^{-1} X'y$, $\nu = T-p$ and $s^2 = \frac{1}{\nu}(y - X\hat{\beta})'(y-X\hat{\beta})$.

Using Bayes' theorem, the likelihood function in (2.3) is combined with a prior distribution $p(\beta,\sigma)$ of the parameters β and σ to yield a joint posterior distribution $p(\beta,\sigma \mid y)$ for these parameters, that is

(2.4) $$p(\beta,\sigma \mid y) = K\, p(\beta,\sigma)\, \ell(\beta,\sigma \mid y)$$

where $$K^{-1} = \int_R p(\beta,\sigma)\, \ell(\beta,\sigma \mid y)\; d\beta\, d\sigma.$$

From the joint posterior distribution of β and σ, we can then derive marginal and conditional posterior distributions for σ and for particular elements of β. Clearly the form of our posterior distribution will depend on the kind of prior information which we have available and the way in which we represent it. In what follows, we consider several formulations which have appeared in the literature and then go on to present and analyze two models which we have developed.

2.2 Locally Uniform Prior Distributions

In problems involving estimation of location and scale parameters, it has been argued in several previous works--Jeffreys (1961), Savage (1961), Box and Tiao (1962)-- that, in many practical situations, it is appropriate to use Bayes' theorem with the assumption that the location parameters and the logarithm of the scale parameters are independent and have locally uniform prior distributions. By a locally uniform prior distribution we mean a distribution function which is practically uniform over the region in which the likelihood function assumes appreciable values, and at no other point is it of sufficiently great magnitude as to become appreciable when multiplied by the likelihood. When such prior distributions are employed, the posterior distribution of the location parameters and the logarithm of the scale parameters is closely approximated by the likelihood

function. In the context of the present problem, since the β_i's are location parameters and σ is a scale parameter, we have then:

$$(2.5a)\qquad p(\beta) \propto k_1$$

$$(2.5b)\qquad p(\log\sigma) \propto k_2 \quad \text{or} \quad p(\sigma) \propto \frac{1}{\sigma}$$

Substituting (2.3) and (2.5) in (2.4), the joint posterior distribution of β and σ is:

$$(2.6)\qquad p(\beta, \sigma|y) = \text{const.}\ \sigma^{-(T+1)} \exp\left\{-\frac{1}{2\sigma^2}[\nu s^2 + Q(\beta, \hat{\beta}, Z)]\right\}.$$

This posterior distribution can be written as

$$p(\beta, \sigma|y) = p(\sigma|y)\, p(\beta|\sigma, y)$$

where

$$(2.7)\qquad p(\sigma|y) = \text{const.}\ \sigma^{-(T-p+1)} \exp\left\{-\frac{\nu s^2}{2\sigma^2}\right\}$$

and

$$(2.8)\qquad p(\beta|\sigma, y) = \text{const.}\ \sigma^{-p} \exp\left\{-\frac{1}{2\sigma^2} Q(\beta, \hat{\beta}, Z)\right\}.$$

We see that (2.7) is in the form of an "inverted" gamma distribution and (2.8) is a multivariate normal distribution with mean $\hat{\beta}$ and covariance matrix $\sigma^2 Z^{-1}$.

When σ is unknown, the marginal posterior distribution of β is obtained by integrating the joint posterior density function over σ, that is,

$$(2.9)\qquad p(\beta|y) = \int_0^\infty p(\beta, \sigma|y)\, d\sigma = \text{const.}\left\{1 + \frac{Q(\beta, \hat{\beta}, Z)}{\nu s^2}\right\}^{-\frac{\nu+p}{2}}.$$

By taking $t_i = (\beta^i - \hat{\beta}^i) / s(z^{ii})^{\frac{1}{2}}$ and $r^{ij} = z_{ij}\,(z^{ii} z^{jj})^{\frac{1}{2}}$, we obtain

$$(2.10)\qquad p(t) = \text{const.}\left\{1 + \frac{\sum_{ij} r^{ij} t_i t_j}{\nu}\right\}^{-\frac{\nu+p}{2}}$$

which is a multivariate t distribution, a result derived by Savage (1961) using the Bayesian approach. It can easily be shown that the marginal

posterior distribution of a subset of the elements of β is also in the same form as in (2.9) and can therefore be transformed into a multivariate t distribution. In addition, the marginal distribution of the quantity t_i is simply a univariate t distribution with T-p degrees of freedom.

We note that these results can also be derived from Fisher's fiducial theory. Further, from the sampling theory point of view, the statistics $\hat{\beta}$ and s are regarded as random variables. The distribution in (2.10) is then precisely the joint distribution of the quantities $t_i = (\hat{\beta}^i - \beta^i) / s(z^{ii})^{\frac{1}{2}}$, $i = 1, 2, \ldots, p$, as shown by Cornish (1954) and by Dunnet and Sobel (1954). There is, of course, nothing new in the above. We record these results as an introduction to the more general models which we present below.

2.3 Normal-Gamma Representation of Prior Distributions

In situations where some prior information about the parameter β is available, we can take as our joint prior distribution for β and a certain scale parameter σ_1:*

$$(2.11) \qquad p(\beta, \sigma_1) = p(\sigma_1)\, p(\beta|\sigma_1)$$

where

$$(2.12) \qquad p(\sigma_1) = \text{const.}\ \sigma_1^{-(\nu_1+1)} \exp\left\{-\frac{\nu_1 s_1^2}{2\sigma_1^2}\right\}$$

and

$$(2.13) \qquad p(\beta|\sigma_1) = \text{const.}\ \sigma_1^{-p} \exp\left\{-\frac{1}{2\sigma_1^2} Q(\beta, \tilde{\beta}, Z_1)\right\}.$$

The quantities ν_1, s_1^2 and the elements of $\tilde{\beta}$ and Z_1 are all known constants; the matrix Z_1 is assumed to be non-negative definite. The prior distribution

---

*The reasons for introducing σ_1 will be made clear in the following discussion where various models are considered.

in (2.11) is called a "normal-gamma" distribution by Raiffa and Schlaifer (1961) and is seen to be in the same form as the posterior distribution of β and σ in (2.6). It can be used, for instance, when experiments are conducted sequentially and the posterior distribution of the parameter(s) of previous experiments are taken as the prior distribution for the current experiment. Suppose the likelihood function of our previous experiments takes the form:

$$(2.14) \qquad \ell(\beta,\ \sigma_1 | y_1) = (\sigma_1 \sqrt{2\pi})^{-T_1} \exp\left\{-\ (1/2\ \sigma_1^2)(y_1 - X_1\beta)'(y_1 - X_1\beta)\right\}.$$

Then, upon making similar assumptions about the prior distributions for β and σ_1 as discussed in Section 2.2, and by setting

$$Z_1 = X_1'X_1,\ \tilde{\beta} = Z_1^{-1}X_1'y_1,\ \nu_1 = T_1 - p \text{ and } s_1^2 = \frac{1}{\nu_1}(y_1 - X_1\tilde{\beta})'(y_1 - X_1\tilde{\beta}),$$

we find that the posterior distribution of β and σ_1 is precisely that given in (2.11).

In taking $p(\beta,\ \sigma_1)$ in (2.11) as the prior distribution to be combined with the likelihood function in (2.2), we immediately see that the exact form of the posterior distribution of β will depend upon our knowledge about the relationship between the scale parameter σ which appears in (2.2) and the new scale parameter σ_1 introduced in (2.11). In what follows we distinguish three different situations: (i) σ_1 is functionally related to σ; (ii) σ_1 is known to take some fixed value σ_{10} and is independent of σ; and (iii) σ_1 is unknown and independent of σ.

2.4 Situation Where σ_1 and σ are Functionally Related

Raiffa and Schlaifer (1961) have considered the case in which σ_1 is proportional to σ with a known factor of proportionality, that is, $\sigma_1 = k\sigma$ with the value of k fixed. Since k is known, there is no loss in generality to assume that $k = 1$ so that $\sigma_1 = \sigma$. This assumption is appropriate, for example, in situations in which experiments are conducted sequentially under well controlled conditions which insure constancy of the variances of random disturbances in all experiments. The prior distribution of β and σ_1 in (2.11), which can be regarded as the posterior distribution of these parameters arising from previous experiments, then provides a priori information for both the parameters β and the scale parameter σ. When this prior distribution is employed in conjunction with the likelihood function in (2.2), the joint posterior distribution of β and σ is given by:

$$(2.15) \qquad p(\beta, \sigma|y) = p(\sigma|y)\, p(\beta|\sigma, \; y)$$

where

$$p(\sigma|y) = \text{const.}\; \sigma^{-(\nu_1+T+1)} \exp\left\{ - \frac{\nu s^2 + \nu_1 s_1^2}{2\sigma^2} \right\}$$

$$p(\beta|\sigma, y) = \text{const.}\; \sigma^{-p} \exp\left\{ - \frac{1}{2\sigma^2}\; Q(\beta, \hat{\hat{\beta}}, Z_2) \right\}$$

$$Z_2 = Z + Z_1 \quad \text{and} \quad \hat{\hat{\beta}} = Z_2^{-1}\,(Z\hat{\beta} + Z_1\,\tilde{\beta}).$$

On integrating out σ from (2.15), we obtain the posterior distribution of β,

$$(2.16) \qquad p(\beta|y) = \text{const.} \left\{ 1 + \frac{Q(\beta, \hat{\hat{\beta}}, Z_2)}{\bar{\nu}\; \bar{s}^2} \right\}^{-\frac{1}{2}(\bar{\nu}+p)}$$

with $\quad \bar{\nu} = \nu_1 + T \quad$ and $\quad \bar{s}^2 = \frac{1}{\bar{\nu}}\,(\nu_1 s_1^2 + \nu s^2).$

This distribution is in the same form as that given in (2.9) and can be transformed into a multivariate t distribution as indicated above.

2.5 Situation in Which σ_1 is Known

In many circumstances, as Theil (1962) has pointed out, the assumption that σ_1 and σ are functionally related is inappropriate. For instance, in econometric analysis it is frequently the case that theoretical considerations may lead the investigator to impose certain, perhaps imprecise, a priori restrictions on the value of β. The conditional prior distribution of β in (2.13) for some assigned value of σ_1, say $\sigma_1 = \sigma_{10}$, may be utilized as a mathematical representation of these a priori restrictions with the assigned σ_{10} measuring, in some sense, the investigator's uncertainty about them.* Since σ_1 is now regarded as a measure of subjective feelings, whereas σ in the likelihood function is a measure of experimental error, there is little reason for supposing that they are functionally related. Thus, assigning the value σ_{10} to σ_1 provides us with no information about σ. We may then follow the analysis in Section 2.2 and take log σ to be locally uniformly distributed a priori. With these assumptions, the posterior distribution of β is:

$$(2.17) \qquad p(\beta|y) = k^{-1} \exp\left\{-\frac{1}{2\sigma_1^2} Q(\beta, \tilde{\beta}, Z_1)\right\} \left\{1 + \frac{Q(\beta, \hat{\beta}, Z)}{\nu s^2}\right\}^{-\frac{\nu+p}{2}}$$

where

$$k = \int_R \exp\left\{-\frac{1}{2\sigma_1^2} Q(\beta, \tilde{\beta}, Z_1)\right\} \left\{1 + \frac{Q(\beta, \hat{\beta}, Z)}{\nu s^2}\right\}^{-\frac{\nu+p}{2}} d\beta \; .$$

This posterior distribution is seen to be in the form of the product of a multivariate normal distribution and a multivariate t distribution. Hereafter, we shall denote a distribution of this type as a multivariate "normal-t" distribution. We note that, when ν tends to infinity, the expression $\left\{1 + \frac{Q(\beta, \hat{\beta}, Z)}{\nu s^2}\right\}^{-\frac{\nu+p}{2}}$ tends to:

---

*In addition to assigning a value to σ_1, it is of course necessary to assign values to $\tilde{\beta}$ and the matrix Z_1 in (2.13).

$$(2.18)\quad \lim_{\nu\to\infty} \left\{1 + \frac{Q(\beta, \hat{\beta}, Z)}{\nu s^2}\right\}^{-\frac{\nu+p}{2}} = \exp\left\{-\frac{1}{2s^2}\; Q(\beta, \hat{\beta}, Z)\right\} .$$

Thus, in the limit, we have for the posterior distribution of β in (2.17),

$$(2.19)\quad \lim_{\nu\to\infty} p(\beta|y) = \frac{|C|^{\frac{1}{2}}}{(2\pi)^{\frac{p}{2}}} \exp\left\{-\frac{1}{2} Q(\beta, \tilde{\tilde{\beta}}, C)\right\}$$

with $C = \frac{1}{\sigma_1^2} Z_1 + \frac{1}{s^2} Z$ and $\tilde{\tilde{\beta}} = C^{-1} \left(\frac{1}{\sigma_1^2} Z_1 \tilde{\beta} + \frac{1}{s^2} Z \hat{\beta}\right)$

which is a multivariate normal distribution with mean $\tilde{\tilde{\beta}}$ and covariance matrix C^{-1}. For finite values of ν, the normalizing constant k in (2.17) is a p-dimensional integral which cannot be expressed in terms of simple functions. Nevertheless, it can be approximated using methods similar to those described in Section 3.

Before leaving this section, we shall make a few remarks about the work of Theil (1962) and Theil and Goldberger (1960) in connection with the use of prior knowledge in regression analysis. Theil and Goldberger are primarily interested in utilizing prior information about β in conjunction with a sample to provide a point estimate of β which incorporates both prior and sample information. In their treatment, the regression model is specified as that given in (2.1) except for the normality assumption, that is

$$y = X\beta + \epsilon$$

with $\quad E(\epsilon) = 0 \quad$ and $\quad E(\epsilon\epsilon') = I\sigma^2$.

The prior information about β can be put in the form:

$$(2.20)\qquad y_1 = X_1\,\beta + \epsilon_1$$

where the elements of ϵ_1 are independently distributed, each with zero mean and <u>known</u> variance σ_1^2. Further, σ_1^2 is assumed to be functionally

independent of σ^2. From the sampling theory point of view, they show that, when σ^2 is known the statistic

$$(2.21)\quad \check{\beta} = \left(\frac{1}{\sigma^2} X'X + \frac{1}{\sigma_1^2} X_1'X_1\right)^{-1} \left(\frac{1}{\sigma^2} X'y + \frac{1}{\sigma_1^2} X_1'y_1\right),$$

is the minimum variance linear unbiased estimator. In case σ^2 is not known, Theil substitutes s^2, the sample variance, for σ^2 in (2.21) and proceeds to show that the resulting statistic $\approx\beta$, given by

$$(2.22)\quad \overset{\approx}{\beta} = \left(\frac{1}{s^2} X'X + \frac{1}{\sigma_1^2} X_1'X_1\right)^{-1} \left(\frac{1}{s^2} X'y + \frac{1}{\sigma_1^2} X_1'y_1\right),$$

differs from $\check{\beta}$ by a quantity which is of order T^{-1} in probability.

It may be of interest to observe the parallelism of the above results and those from the Bayesian formulation we have considered. When a normality assumption is added, the likelihood function corresponding to (2.20) is proportional to the expression given in (2.14). For the case σ^2 known, it can readily be shown that the posterior distribution of β is multivariate normal with mean given by the expression in (2.21). In the case where σ^2 is not known, the expression in (2.22) is precisely the limiting mean for the multivariate "normal-t" distribution as ν tends to infinity [see equation (2.19)]. This result is, of course, to be expected since, except for the normality assumption about the disturbances, all other underlying assumptions are very much the same in both approaches.

2.6 Situation in Which σ_1 is Regarded as a Variable Parameter and Independent of σ

We have considered two models above, one in which it is assumed that $\sigma_1 = k\sigma$ with k known, and the other in which σ_1 is independent of σ but takes on a fixed value σ_{10}. As a generalization of the second model, we now consider σ and σ_1 to be independent variable parameters.

This formulation will be applicable, for example, in the following situation. Suppose that the results of two sets of experiments are utilized to make inferences about β and that the associated likelihood functions are given by $\ell(\beta, \sigma_1|y_1)$ in (2.14) and $\ell(\beta, \sigma|y)$ in (2.2), respectively. Suppose further that these two sets of experiments are carried out under quite different conditions so that there is no basis for assuming any relationship between σ_1 and σ. Follcwing the discussion in section 2.3 , it seems appropriate to take the normal-gamma distribution $p(\beta, \sigma_1)$ in (2.11) as the posterior distribution associated with the first set of experiments (see discussion in Section 2.3). This distribution can then be regarded as representing prior information about β and σ_1 for the analysis of the second set. Since σ_1 and σ are independent, information about σ_1 represented by the marginal distribution $p(\sigma_1)$ in (2.12) contributes nothing to the investigator's knowledge about σ. Thus, all that is of interest in $p(\beta, \sigma_1)$ is the information concerning β. This is, of course, represented by the marginal distribution $p(\beta)$, namely

$$(2.23) \qquad p(\beta) = \int_0^\infty p(\beta, \sigma_1)\, d\sigma_1$$

$$= \text{const.} \left\{1 + \frac{Q(\beta, \tilde{\beta}, Z_1)}{\nu_1 s_1^2}\right\}^{-\frac{\nu_1+p}{2}} .$$

Using (2.23) as the prior distribution of β and upon making the same assumption about the prior distribution of σ as in Section 2.2, the posterior distribution of β is readily found to be:

$$(2.24)\quad p(\beta|y) = k_1^{-1}\left\{1 + \frac{Q(\beta,\, \tilde{\beta},\, Z_1)}{\nu_1\, s_1^2}\right\}^{-\frac{\nu_1+p}{2}}\left\{1 + \frac{Q(\beta,\, \hat{\beta},\, Z)}{\nu s^2}\right\}^{-\frac{\nu+p}{2}}$$

with

$$k_1 = \int_R \left\{1 + \frac{Q(\beta,\, \tilde{\beta},\, Z_1)}{\nu_1\, s_1^2}\right\}^{-\frac{\nu_1+p}{2}}\left\{1 + \frac{Q(\beta,\, \hat{\beta},\, Z)}{\nu s^2}\right\}^{-\frac{\nu+p}{2}} d\beta\ .$$

This distribution is in the form of the product of two quantities each of which can be transformed into a multivariate t distribution. Hereafter we shall denote a distribution of this kind as a multivariate "double-t" distribution. As in the case of a multivariate "normal-t" distribution, the normalizing constant k_1 in (2.24) is a p-dimensional integral. This may lead to certain practical difficulties in the numerical evaluation of the posterior distribution, particularly when p is large. Similar difficulties will also be encountered if one is interested in making inferences about a subset of the elements of β, since in this case it does not appear possible to express the corresponding marginal posterior distribution of the subset of interest in terms of simple functions. In the following section, we develop a method by which both the posterior distribution in (2.24) and the marginal distributions of elements of β can be approximated.

We note that when the vector β has only one element ($p = 1$) and the elements of the corresponding (Tx1) matrix X in (2.2) and (T_1x1) matrix X_1 in (2.14) have the same value, unity, the posterior distribution in (2.24) takes the following form:

$$(2.25)\qquad p(\beta|y) = k^{-1}\left\{1 + \frac{(\nu_1+1)(\beta-\bar{y}_1)^2}{\nu_1\, s_1^2}\right\}^{-\frac{\nu_1+1}{2}} \left\{1 + \frac{(\nu+1)(\beta-\bar{y})^2}{\nu s^2}\right\}^{-\frac{\nu+1}{2}}$$

where
$$k = \int_{-\infty}^{\infty}\left\{1 + \frac{(\nu_1+1)(\beta-\bar{y}_1)^2}{\nu_1\, s_1^2}\right\}^{-\frac{\nu_1+1}{2}}\left\{1 + \frac{(\nu+1)(\beta-\bar{y})^2}{\nu s^2}\right\}^{-\frac{\nu+1}{2}} d\beta$$

and the quantities $\bar{y}_1$, $\bar{y}$, s_1^2 and s^2 are respectively, the sample means and sample variances for the two sets of experiments. This result corresponds to the problem of making inferences about a population mean when samples are drawn from two normal populations with common mean and unequal variances. It is of interest to note that the distribution given in (2.25) is exactly the same as that obtained by Fisher (1961a, 1961b) from the fiducial theory point of view. He proceeded to expand this distribution in an asymptotic series in powers of ν_1 and ν, from which probability integrals of β can be approximated. We may remark here that our development in Section 3 closely parallels Fisher's procedure.

It is easy to see that the analysis in this section can be immediately generalized to cover situations in which several sets of experiments are conducted sequentially (or concurrently) but under quite different conditions. Suppose that the likelihood function for the i^{th} set of experiments can be represented by:

$$(2.26)\qquad \ell(\beta,\ \sigma_i|y_i) = (\sigma_i\sqrt{2\pi})^{-T_i}\ \exp\left\{-\left(\frac{1}{2\sigma_i^2}\right)(y_i-X_i\beta)'(y_i-X_i\beta)\right\}$$

where $i = 1, 2, \ldots, K$ say. Then, by taking the σ_i's as independent scale parameters we obtain the following posterior distribution of β:

$$(2.25)\qquad p(\beta|y) = k^{-1}\left\{1 + \frac{(\nu_1+1)(\beta-\bar{y}_1)^2}{\nu_1\, s_1^2}\right\}^{-\frac{\nu_1+1}{2}} \left\{1 + \frac{(\nu+1)(\beta-\bar{y})^2}{\nu s^2}\right\}^{-\frac{\nu+1}{2}}$$

where
$$k = \int_{-\infty}^{\infty}\left\{1 + \frac{(\nu_1+1)(\beta-\bar{y}_1)^2}{\nu_1\, s_1^2}\right\}^{-\frac{\nu_1+1}{2}} \left\{1 + \frac{(\nu+1)(\beta-\bar{y})^2}{\nu s^2}\right\}^{-\frac{\nu+1}{2}} d\beta$$

and the quantities $\bar{y}_1$, $\bar{y}$, s_1^2 and s^2 are respectively, the sample means and sample variances for the two sets of experiments. This result corresponds to the problem of making inferences about a population mean when samples are drawn from two normal populations with common mean and unequal variances. It is of interest to note that the distribution given in (2.25) is exactly the same as that obtained by Fisher (1961a, 1961b) from the fiducial theory point of view. He proceeded to expand this distribution in an asymptotic series in powers of ν_1 and ν, from which probability integrals of β can be approximated. We may remark here that our development in Section 3 closely parallels Fisher's procedure.

It is easy to see that the analysis in this section can be immediately generalized to cover situations in which several sets of experiments are conducted sequentially (or concurrently) but under quite different conditions. Suppose that the likelihood function for the i^{th} set of experiments can be represented by:

$$(2.26)\qquad \ell(\beta,\, \sigma_i|y_i) = (\sigma_i\sqrt{2\pi})^{-T_i}\ \exp\left\{-\left(\frac{1}{2\sigma_i^2}\right)(y_i-X_i\beta)'(y_i-X_i\beta)\right\}$$

where $i = 1, 2, \ldots, K$ say. Then, by taking the σ_i's as independent scale parameters we obtain the following posterior distribution of β:

$$(2.27) \qquad p(\beta|y) = \omega \prod_{i=1}^{K} \left\{1 + \frac{Q(\beta,\, \hat{\beta}_i,\, Z_i)}{\nu_i\, s_i^2}\right\}^{-\frac{\nu_i+p}{2}}$$

with

$$\omega^{-1} = \int_R \prod_{i=1}^{K} \left\{1 + \frac{Q(\beta,\, \hat{\beta}_i,\, Z_i)}{\nu_i\, s_i^2}\right\}^{-\frac{\nu_i+p}{2}} d\beta$$

$$\nu_i = T_i - p \qquad Z_i = X_i' X_i$$

$$\hat{\beta}_i = Z_i^{-1} X_i' y_i \qquad \text{and } s_i^2 = \frac{1}{\nu_i} (y_i - X_i'\hat{\beta}_i)'(y_i - X_i'\hat{\beta}_i).$$

This distribution is seen to be the product of K quantities each of which can be expressed as a multivariate t distribution. It may, therefore, be denoted as a multivariate "multiple-t" distribution and can be approximated numerically using methods similar to those described in the next section.

III. ASYMPTOTIC EXPRESSION FOR THE MULTIVARIATE "DOUBLE-t" POSTERIOR DISTRIBUTION

3.1 The Joint Posterior Distribution

In the preceding section, we have shown that, when σ_1 and σ are regarded as independent variable parameters, the corresponding posterior distribution of β is in the form of the product of two multivariate t distributions. [See (2.24).] The normalizing constant is a p-dimensional integral which is in general difficult to evaluate even on a fast computer, especially when p is large. Nevertheless, we now show that, by expanding the posterior distribution into an asymptotic series in powers of ν^{-1} and ν_1^{-1}, we can reduce the problem of integration to a problem of evaluating the mixed moments of two quadratic forms. The same procedure is then applied in the next section to obtain an asymptotic expression for the marginal posterior distributions of elements of β.

Since s_1^2 and s^2 in (2.24) are known quantities, they can be suppressed by setting

$$M = \frac{1}{s_1^2} Z_1 \qquad \text{and} \quad B = \frac{1}{s^2} Z.$$

We can then write (2.24) as:

$$(3.1) \qquad p(\beta|y) = k_1^{-1} \left\{1 + \frac{Q(\beta, \tilde{\beta}, M)}{\nu_1}\right\}^{-\frac{\nu_1+p}{2}} \left\{1 + \frac{Q(\beta, \hat{\beta}, B)}{\nu}\right\}^{-\frac{\nu+p}{2}}$$

with

$$k_1 = \int_R \left\{1 + \frac{Q(\beta, \tilde{\beta}, M)}{\nu_1}\right\}^{-\frac{\nu_1+p}{2}} \left\{1 + \frac{Q(\beta,\hat{\beta}, B)}{\nu}\right\}^{-\frac{\nu+p}{2}} d\beta.$$

The expression $\left\{1 + \frac{Q(\beta, \hat{\beta}, B)}{\nu}\right\}^{-\frac{\nu+p}{2}}$ can be written:

$$\left\{1 + \frac{Q(\beta, \hat{\beta}, B)}{\nu}\right\}^{-\frac{\nu+p}{2}} = \exp\left\{-\tfrac{1}{2} Q(\beta, \hat{\beta}, B)\right\} \cdot \exp\left\{\tfrac{1}{2} Q(\beta, \hat{\beta}, B) - \frac{\nu+p}{2} \log\left[1 + \frac{Q(\beta, \hat{\beta}, B)}{\nu}\right]\right\}.$$

Expanding the second factor on the right in powers of ν^{-1}, we obtain:

$$(3.2) \qquad \left\{1 + \frac{Q(\beta, \hat{\beta}, B)}{\nu}\right\}^{-\frac{\nu+p}{2}} = \exp\left\{-\tfrac{1}{2} Q(\beta, \hat{\beta}, B)\right\} \sum_{i=0}^{\infty} P_i \, \nu^{-i}$$

where

$$P_0 = 1$$

$$P_1 = \frac{1}{4}\left[Q^2(\beta, \hat{\beta}, B) - 2p\, Q(\beta, \hat{\beta}, B)\right]$$

$$P_2 = \frac{1}{96}\left[3Q^4(\beta, \hat{\beta}, B) - 4(3p+4)\, Q^3(\beta, \hat{\beta}, B) + 12p\,(p+2)\, Q^2(\beta,\hat{\beta},B)\right]$$

. . . .

Similarly, we have that

$$(3.3)\quad \left\{1 + \frac{Q(\beta,\ \tilde{\beta},\ M)}{\nu_1}\right\}^{-\frac{\nu_1+p}{2}} = \exp\left\{-\tfrac{1}{2}\, Q(\beta,\ \tilde{\beta},\ M)\right\} \sum_{i=0}^{\infty} q_i\ \nu_1^{-i}$$

where

$$q_o = 1$$

$$q_1 = \frac{1}{4}\ [\ Q^2\ (\beta,\tilde{\beta},M)\ -\ 2p\ Q(\beta,\tilde{\beta},M)]$$

$$q_2 = \frac{1}{96}\,[\,3Q^4\ (\beta,\tilde{\beta},M)\ -\ 4(3p+4)\ Q^3\ (\beta,\tilde{\beta},M)\ +\ 12p\ (p+2)\ Q^2\ (\beta,\tilde{\beta},M)]$$

. . . .

Substituting (3.2) and (3.3) into (3.1) and after a little reduction, we can express the posterior distribution as:

$$(3.4)\quad p(\beta|y) = \frac{1}{W}\,\frac{|D|^{\frac{1}{2}}}{(2\pi)^{p/2}}\ \exp\left\{-\tfrac{1}{2}\, Q(\beta,\ \bar{\beta},\ D)\right\} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_i\ q_j\ \nu^{-i}\ \nu_1^{-j}$$

where $D = B + M,\quad \bar{\beta} = D^{-1}\ (B\hat{\beta}\ +\ M\tilde{\beta})$

and

$$(3.5)\quad W = \int_{R} \frac{|D|^{\frac{1}{2}}}{(2\pi)^{p/2}}\ \exp\left\{-\tfrac{1}{2}\, Q(\beta,\ \bar{\beta},D)\right\} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p_i\ q_j\ \nu^{-i}\ \nu_1^{-j}\ \ d\beta\ .$$

The integral W in (3.5) can be integrated term by term. From (3.2) and (3.3), we see that each term is, in fact, a bivariate polynomial in the mixed moments of the quadratic forms $Q(\beta,\ \hat{\beta},\ B)$ and $Q(\beta,\ \tilde{\beta},\ M)$ where the variables β have a multivariate normal distribution with mean $\bar{\beta}$ and covariance matrix D^{-1}. For this problem, it appears much simpler to obtain the mixed moments indirectly by first finding the mixed cumulants. It is straightforward to verify that the joint cumulant generating function of $Q(\beta,\ \tilde{\beta},\ M)$ and $Q(\beta,\ \hat{\beta},\ B)$ is

(3.6) $$\kappa(t_1, t_2) = \log \int_R \frac{|D|^{\frac{1}{2}}}{(2\pi)^{p/2}} \exp\left\{t_1\, Q(\beta,\hat{\beta},B) + t_2\, Q(\beta,\tilde{\beta},M) - \tfrac{1}{2}\, Q(\beta,\bar{\beta},D)\right\} d\beta$$

$$= -\tfrac{1}{2} \log \left|I - 2D^{-1}(t_1 B + t_2 M)\right| + t_1\, \eta_1'\, B\, \eta_1 + t_2\, \eta_2'\, M\, \eta_2 + 2(t_1 B \eta_1 + t_2 M \eta_2)'\,(D - 2t_1 B - 2t_2 M)^{-1}\,(t_1 B \eta_1 + t_2 M \eta_2)$$

where $\eta_1 = \bar{\beta} - \hat{\beta}$ and $\eta_2 = \bar{\beta} - \tilde{\beta}$.

Upon differentiating (3.6) and after some algebraic reduction, we find: (see Appendix)

(3.7) $$\kappa_{10} = \text{tr. } D^{-1} B + \eta_1'\, B\, \eta_1$$

$$\kappa_{01} = \text{tr. } D^{-1} M + \eta_2'\, M\, \eta_2$$

$$\kappa_{rs} = 2^{r+s-1}(r+s-2)!\left\{(r+s-1)\,\text{tr. } D^{-1} G^{rs} + (r\eta_1 + s\eta_2)'\, G^{rs}\,(r\eta_1 + s\eta_2) - r\eta_1'\, G^{rs}\, \eta_1 - s\eta_2'\, G^{rs}\, \eta_2\right\} \qquad r + s \geq 2$$

where $G^{rs} = D(D^{-1} B)^r (D^{-1} M)^s$.

Employing the bivariate moment-cumulant inversion formulae as given by Cook (1951), the integral in (3.5) can be written as

(3.8) $$W = \sum_{i=o}^{\infty} \sum_{j=o}^{\infty} b_{ij}\, \nu^{-i}\, \nu_1^{-j}$$

where $b_{oo} = 1$

$$b_{10} = \frac{1}{4}\,[\kappa_{20} + \kappa_{10}^2 - 2p\, \kappa_{10}]$$

$$b_{01} = \frac{1}{4}\,[\kappa_{02} + \kappa_{01}^2 - 2p\, \kappa_{01}]$$

$$b_{11} = \frac{1}{16}[\kappa_{22}+\kappa_{20}\kappa_{02}+2\kappa_{11}^2+4\kappa_{11}\kappa_{01}\kappa_{10}+\kappa_{10}^2\kappa_{01}^2+2\kappa_{21}\kappa_{01}+2\kappa_{12}\kappa_{10}$$
$$+\kappa_{20}\kappa_{01}^2+\kappa_{02}\kappa_{10}^2-2p(\kappa_{12}+\kappa_{21}+\kappa_{02}\kappa_{10}+\kappa_{20}\kappa_{01}+2\kappa_{11}\kappa_{10}$$
$$+2\kappa_{11}\kappa_{01}+\kappa_{10}\kappa_{01}^2+\kappa_{01}\kappa_{10}^2)+4p^2(\kappa_{11}-\kappa_{01}\kappa_{10})]$$

$$b_{20} = \frac{1}{96}[3(\kappa_{40}+3\kappa_{20}^2+4\kappa_{30}\kappa_{10}+6\kappa_{20}\kappa_{10}^2+\kappa_{10}^4)-4(3p+4)$$
$$(\kappa_{30}+3\kappa_{20}\kappa_{10}+\kappa_{10}^3)+12p(p+2)(\kappa_{20}+\kappa_{10}^2)]$$

$$b_{02} = \frac{1}{96}[3(\kappa_{04}+3\kappa_{02}^2+4\kappa_{03}\kappa_{01}+6\kappa_{02}\kappa_{01}^2+\kappa_{01}^4)$$
$$-4(3p+4)(\kappa_{03}+3\kappa_{02}\kappa_{01}+\kappa_{01}^3)+12p(p+2)(\kappa_{02}+\kappa_{01}^2)]$$

. . . .

Substituting the results in (3.8) into (3.4), we obtain the following asymptotic expression for the posterior distribution of β:

$$(3.9)\qquad p(\beta|y) = \frac{|D|^{\frac{1}{2}}}{(2\pi)^{p/2}} \exp\left\{-\tfrac{1}{2}Q(\beta,\bar{\beta},D)\right\}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} d_{ij}\,\nu^{-i}\,\nu_1^{-j}$$

where $d_{00} = 1$

$$d_{10} = p_1 - b_{10}$$
$$d_{01} = q_1 - b_{01}$$
$$d_{11} = (p_1-b_{10})(q_1-b_{01}) + b_{10}\,b_{01} - b_{11}$$
$$d_{20} = p_2 - b_{20} - p_1\,b_{10} + b_{10}^2$$
$$d_{02} = q_2 - b_{02} - q_1\,b_{01} + b_{01}^2$$

. . . .

Expressions for additional terms d_{12}, d_{21}, d_{22}, etc. can similarly be found if desired.

The posterior distribution is thus expressed in the form of a multivariate normal distribution multiplied by a power series in ν^{-1} and ν_1^{-1}.

When both ν and ν_1 tend to infinity, all terms of the power series except the leading one vanish so that, in the limit, the posterior distribution is multivariate normal with mean $\bar{\beta}$ and covariance matrix D^{-1}.* For finite values of ν and ν_1, the terms in the power series can be regarded as "corrections" in a normal approximation to the multivariate "double-t" distribution. From (3.2), (3.3) and (3.7), we see that numerical evaluation of the coefficients in the power series involves merely matrix inversions and multiplications, operations which are easily performed on an electronic computer.

We note that when the posterior distribution is a univariate distribution as in (2.25), the results in (3.9) are in exact agreement with those obtained by Fisher (1961b) in a similar treatment of the problem (see discussion in Section 2.6). In Fisher's derivation, each term of the integral W in (3.5) was expressed in terms of the moments of a univariate normal distribution. It can therefore be evaluated directly without making use of the mixed-cumulant formulae given in (3.7) which seem more convenient for the multivariate case considered here.

For the univariate case, posterior pr abilities can be calculated using the formulae given in Fisher's paper cited above. When $p > 1$, numerical evaluation of joint probabilities becomes exceedingly cumbersome. Nevertheless, using the expression (3.9) the density function can be calculated conveniently. When $p = 2$, the joint distribution contours can of course be plotted, giving the investigator a complete summary of the information about β. This will be illustrated by an example in Section 4.

---

*It should be obvious that if one of the ν and ν_1 tends to infinity while the other remains finite, the multivariate "double-t" posterior distribution tends to the multivariate "normal-t" form. Our above development can easily be modified to yield an asymptotic expression for the latter distribution.

3.2 The Marginal Posterior Distribution

When interest centers on a subset of the elements of β, say $\beta_1 = (\beta^1, \ldots, \beta^\ell)$, an asymptotic expression for the corresponding marginal posterior distribution can be obtained by integrating out the remaining elements, $\beta_2 = (\beta^{\ell+1}, \ldots, \beta^p)$ from the joint distribution in (3.9). We have that

$$\text{(3.10)}\quad p(\beta_1|y) = \frac{|D|^{\frac{1}{2}}}{(2\pi)^{p/2}} \int_{R'} \exp\left\{-\tfrac{1}{2}(Q, \bar{\beta}, D)\right\} \sum_{i=o}^{\infty} \sum_{j=o}^{\infty} d_{ij}\, \nu^{-i}\, \nu_1^{-j}\, d\beta_2$$

Denoting $\bar{\beta} = (\bar{\beta}_1 \vdots \bar{\beta}_2)$ and partitioning the matrices D and D^{-1} into:

$$D = \begin{array}{cc} \begin{array}{cc} \ell & p-\ell \end{array} & \\ \left|\begin{array}{c:c} D_{11} & D_{12} \\ \hdashline D_{21} & D_{22} \end{array}\right| & \begin{array}{c} \ell \\ p-\ell \end{array} \end{array} \qquad D^{-1} = \begin{array}{cc} \begin{array}{cc} \ell & p-\ell \end{array} & \\ \left|\begin{array}{c:c} V_{11} & V_{12} \\ \hdashline V_{21} & V_{22} \end{array}\right| & \begin{array}{c} \ell \\ p-\ell \end{array} \end{array} ,$$

we can write the marginal posterior distribution as:

$$\text{(3.11)}\quad p(\beta_1|y) = \frac{|V_{11}^{-1}|^{\frac{1}{2}}}{(2\pi)^{\ell/2}} \exp\left\{-\tfrac{1}{2} Q(\beta_1, \bar{\beta}_1, V_{11}^{-1})\right\} f(\beta_1|y)$$

where

$$\text{(3.12)}\quad f(\beta_1|y) = \frac{|D_{22}|^{\frac{1}{2}}}{(2\pi)^{(p-\ell)/2}} \int_{R'} \exp\left\{-\tfrac{1}{2} Q(\beta_2, \alpha, D_{22})\right\} \sum_{i=o}^{\infty} \sum_{j=o}^{\infty} d_{ij}\, \nu^{-i}\, \nu_1^{-j}\, d\beta_2$$

with $\quad \alpha = \bar{\beta}_2 - D_{22}^{-1} D_{21} (\beta_1 - \bar{\beta}_1)$.

From the expressions for d_{ij} given in (3.9), we see that each term in the integral $f(\beta_1|y)$ is a bivariate polynomial in the quadratic forms $Q(\beta, \hat{\beta}, B)$ and $Q(\beta, \tilde{\beta}, M)$ where β_1 is considered fixed and β_2 has a multivariate normal distribution with mean α and covariance matrix D_{22}^{-1}. Adopting the same procedure as that described in the preceding section, and by setting

$$\hat{\beta} = (\hat{\beta}_1 \vdots \hat{\beta}_2), \qquad \tilde{\beta} = (\tilde{\beta}_1 \vdots \tilde{\beta}_2),$$

$$B = \begin{array}{cc} \ell & p-\ell \\ \left[\begin{array}{c:c} B_{11} & B_{12} \\ \hdashline B_{21} & B_{22} \end{array}\right] \end{array} \begin{array}{l} \ell \\ p-\ell \end{array}, \qquad B^{-1} = \begin{array}{cc} \ell & p-\ell \\ \left[\begin{array}{c:c} E_{11} & E_{12} \\ \hdashline E_{21} & E_{22} \end{array}\right] \end{array} \begin{array}{l} \ell \\ p-\ell \end{array},$$

$$M = \begin{array}{cc} \ell & p-\ell \\ \left[\begin{array}{c:c} M_{11} & M_{12} \\ \hdashline M_{21} & M_{22} \end{array}\right] \end{array} \begin{array}{l} \ell \\ p-\ell \end{array}, \qquad M^{-1} = \begin{array}{cc} \ell & p-\ell \\ \left[\begin{array}{c:c} N_{11} & N_{12} \\ \hdashline N_{21} & N_{22} \end{array}\right] \end{array} \begin{array}{l} \ell \\ p-\ell \end{array},$$

we obtain, for the mixed cumulants of $Q(\beta, \hat{\beta}, B)$ and $Q(\beta, \tilde{\beta}, M)$:

$$(3.13) \qquad \omega_{10} = \text{tr. } D_{22}^{-1} B_{22} + \gamma_1' B_{22} \gamma_1 + Q(\beta_1, \hat{\beta}_1, E_{11}^{-1})$$

$$\omega_{01} = \text{tr. } D_{22}^{-1} M_{22} + \gamma_2' M_{22} \gamma_2 + Q(\beta_1, \tilde{\beta}_1, N_{11}^{-1})$$

$$\omega_{rs} = 2^{r+s-1} (r+s-2)! \Big\{ (r+s-1) \text{ tr. } D_{22}^{-1} H^{rs} + (r\gamma_1 + s\gamma_2)' H^{rs} (r\gamma_1 + s\gamma_2) - r\gamma_1' H^{rs} \gamma_1 + s\gamma_2' H^{rs} \gamma_2 \Big\} \qquad r + s \geq 2$$

where

$$H^{rs} = D_{22} (D_{22}^{-1} B_{22})^r (D_{22}^{-1} M_{22})^s$$

$$\gamma_1 = \alpha - \hat{\beta}_2 + B_{22}^{-1} B_{21} (\beta_1 - \hat{\beta}_1)$$

$$\gamma_2 = \alpha - \tilde{\beta}_2 + M_{22}^{-1} M_{21} (\beta_1 - \tilde{\beta}_1) .$$

Using the resultsin (3.13), we can express the marginal posterior distribution of β_1 as:

$$(3.14) \quad p(\beta_1 | y) = \frac{|V_{11}^{-1}|^{\frac{1}{2}}}{(2\pi)^{\ell/2}} \exp \left\{ -\tfrac{1}{2} Q(\beta, \bar{\beta}, V_{11}^{-1}) \right\} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \delta_{ij} \, \nu^{-i} \, \nu_1^{-j}$$

where

$$\delta_{00} = 1$$
$$\delta_{10} = g_{10} - b_{10}$$
$$\delta_{01} = g_{01} - b_{01}$$
$$\delta_{11} = g_{11} - b_{11} - g_{10}\, b_{01} - g_{01}\, b_{10} + 2\, b_{01}\, b_{10}$$
$$\delta_{20} = g_{20} - b_{20} - g_{10}\, b_{10} + b_{10}^2$$
$$\delta_{02} = g_{02} - b_{02} - g_{01}\, b_{01} + b_{01}^2$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$

and the quantities g_{ij} are functions of the mixed cumulants ω_{ij} with the functional relationships exactly the same as those between b_{ij} and κ_{ij} shown in (3.8).

It will be noted that when β_1 consists of only one variable ($\ell = 1$), the quantities δ_{ij} in (3.14) are simply polynomials in that variable. Employing the well known expression for the moments of a normal variable, one can easily derive an asymptotic expression for the moments of β_1. In addition, probability integrals can also be approximated using methods given in Fisher's previously cited paper (1961b).

IV. AN ILLUSTRATIVE EXAMPLE

To illustrate application of the techniques developed in Sections 2 and 3, we analyze a very simple econometric investment model with annual time series data, 1935-1954, relating to two large corporations, General Electric and Westinghouse.* In this model, price deflated gross investment is assumed to be a linear function of expected profitability and beginning of year real capital stock. Following Grunfeld (1958), the value of outstanding shares at the beginning of the year is taken as a measure of a firm's expected profitability. The two investment relations are:

$$
\begin{aligned}
y_1(t) &= \alpha_1 + \beta_1 x_{11}(t) + \beta_2 x_{12}(t) + \epsilon_1(t) \\
y_2(t) &= \alpha_2 + \beta_1 x_{21}(t) + \beta_2 x_{22}(t) + \epsilon_2(t)
\end{aligned}
\tag{4.1}
$$

where t in parentheses denotes the value of a variable in year t, t = 1, 2, ..., 20, and

Variable	General Electric	Westinghouse
Annual real gross investment	$y_1(t)$	$y_2(t)$
Value of shares at beginning of year	$x_{11}(t)$	$x_{21}(t)$
Real capital stock at beginning of year	$x_{12}(t)$	$x_{22}(t)$
Error term	$\epsilon_1(t)$	$\epsilon_2(t)$

The parameters β_1 and β_2 in (4.1) are taken to be the same for the two firms; however, α_1 and α_2 are assumed to be different to allow for certain possible differences in the investment behavior of the two firms. Further, $\epsilon_1(t)$ and $\epsilon_2(t)$ are assumed to be independently and normally distributed

---

*The data are taken from Boot and deWitt (1960).

for all t with zero means and variances σ_1^2 and σ^2, respectively. Since we have no information from which to posit a relationship connecting σ_1^2 and σ^2, we take them to be independent parameters and pursue the development described in Section 2.6.

In the present instance we can regard either General Electric's or Westinghouse's data as being generated "first" and derive a joint posterior distribution of the relevant parameters. This can then serve to represent prior information in the analysis of the second set of data. Or, with locally uniform prior distributions for the parameters in both equations, one can analyze both sets of data at the same time. In both cases the final result is the same joint posterior distribution for α_1, α_2, β_1 and β_2 which is in the form of the product of two multivariate t distributions. On integrating out α_1 and α_2,* the coefficients β_1 and β_2 will be jointly distributed in a bivariate "double t" form. [See equation (2.24).] Numerical values for quantities appearing in (2.24) and (3.9) are shown below:

General Electric	Westinghouse
$\tilde{\beta}_1 = 0.02655$	$\hat{\beta}_1 = 0.05289$
$\tilde{\beta}_2 = 0.1517$	$\hat{\beta}_2 = 0.09241$
$s_1^2 = 777.4463$	$s^2 = 104.3079$
$\nu_1 = 17$	$\nu = 17$

$$M = \begin{bmatrix} 4185.1054 & 299.6748 \\ 299.6748 & 1535.0640 \end{bmatrix} \qquad B = \begin{bmatrix} 9010.5868 & 1871.1079 \\ 1871.1079 & 706.3320 \end{bmatrix}$$

$$\bar{\beta} = (.0373, \quad .1446)$$

---

*If one is interested in the parameters α_1 and α_2, it should be obvious that, a posteriori, they are distributed in the form of two independent t variables. In particular, the difference, $\alpha_1 - \alpha_2$, has the Behrens-Fisher distribution.

A plot of the contours of the joint density surface is shown in Figure 1 along with lines showing the loci of conditional modes. These contours summarize all the relevant information about the coefficients β_1 and β_2. We see that the posterior distribution is concentrated rather sharply in the region $.0278 < \beta_1 < .0468$ and $.1216 < \beta_2 < .1676$, with mode at about (.0373, .1446). Further, β_1 and β_2 are seen to be negatively correlated and the contours are approximately elliptical. The latter is the case because the joint density function is close to its limiting bivariate normal distribution. This arises from the fact that in this example both ν_1 and ν are rather large.

When interest centers on only one of the parameters, say β_1, the expression in (3.14) can be employed to calculate the corresponding marginal distribution. For this example we evaluated (3.14) disregarding terms for which $i + j > 2$. The results are shown by the solid curve in Figure 2. The broken curve in the same figure represents the limiting normal density function with mean $\bar{\beta}_1 = .0373$ and variance $v_{11} = 9.01445 \times 10^{-5}$. It will be noted that the posterior distribution of β_1 is somewhat flatter at the center and fatter in the tails than its limiting distribution. Also, it is slightly skewed. The mean and variance of the distribution of β_1 were computed from (3.14) neglecting terms for which $i + j > 1$. The calculation yielded the following results:

Limiting Normal Distribution	Finite Sample Corrections δ_{10}	δ_{01}	Finite Sample Mean and Variance of β_1
(1)	(2)	(3)	(1) + (2) + (3)
Mean = .0373	$-.000229$	.000191	Mean = .03726
Variance = 9.01445×10^{-5}	$.5985 \times 10^{-5}$	$.0028 \times 10^{-5}$	Variance = 9.6158×10^{-5}

The mean of β_1 is extremely close to its asymptotic value. On the other hand, the variance of β_1 is about 6 percent larger than that of the limiting distribution.

V. SUMMARY

In this paper, we have adopted a Bayesian approach to the problem of integrating prior information into the analysis of the normal regression model. Initially, we reviewed Jeffrey's and Savage's analysis wherein prior knowledge (or lack of substantial prior knowledge) about the regression coefficient β and the logarithm of the scale parameter σ is represented by locally uniform distributions. We then turned to consider a normal-gamma representation of prior information about β and an additional scale parameter σ_1. Here we discussed three possible assumptions about the two scale parameters, namely, (i) $\sigma_1 = \kappa\sigma$ with known value of κ -- the Raiffa and Schlaifer case; (ii) σ_1 fixed and functionally independent of σ; and (iii) both σ_1 and σ unknown and assumed independent a priori.

With assumption (ii), we were able to provide a reinterpretation of the "mixed" estimation procedure of Theil and Goldberger. It was shown that the posterior distribution of β takes the form of a product of multivariate normal and multivariate t distributions.

Under the third assumption, we obtained what may be regarded as a generalization of Fisher's work on the problem of making inferences when samples are drawn from two normal populations with common mean and unequal variances. In this case, it was shown that the posterior distribution of β is in the form of the product of two multivariate t distributions. For computational purposes, the distribution was expanded in an asymptotic series which involved finding the mixed cumulants of pairs of quadratic forms in normal variables. A bivariate example was analyzed in detail.

Appendix

In Section 3.1, we have stated that the joint cumulant generating function of the quadratic forms $Q(\beta, \tilde{\beta}, M)$ and $Q(\beta, \hat{\beta}, B)$ is given by

$$(A.1) \quad \kappa(t_1, t_2) = -\tfrac{1}{2} \log |I-2D^{-1}(t_1B+t_2M)| + t_1\eta_1'B\eta_1 + t_2\eta_2'M\eta_2$$
$$+ 2(t_1B\eta_1 + t_2M\eta_2)' (D-2t_1B-2t_2M)^{-1} (t_1B\eta_1 + t_2M\eta_2).$$

We now derive the expressions for the mixed cumulants shown in (3.7). In our development, we shall make use of the following lemma the proof of which can be found, for example, in Box (1954).

Lemma: Let P be a nxn positive definite symmetric matrix and Q be a nxn nonnegative definite symmetric matrix. Then, for sufficiently small ι, we have

$$\log |I - \iota PQ| = -\sum_{r=1}^{\infty} \frac{\iota^r}{r} \text{ tr. } (PQ)^r.$$

Employing the above lemma and for sufficiently small values of t_1 and t_2, we can expand the first term on the right of (A.1) into:

$$(A.2) \quad -\tfrac{1}{2} \log |I-2D^{-1}(t_1B+t_2M)| = \sum_{r=1}^{\infty} \frac{2^{r-1}}{r} \text{ tr. } (t_1D^{-1}B+t_2D^{-1}M)^r .$$

The quadratic form $t_1\eta_1'B\eta_1$ can be written:

$$(A.3) \quad t_1\eta_1'B\eta_1 = t_1\eta_1'B(D-2t_1B-2t_2M)^{-1} (D-2t_1B-2t_2M) \eta_1$$
$$= t_1\eta_1'B(D-2t_1B-2t_2M)^{-1} D\eta_1 - 2t_1^2\eta_1'B(D-2t_1B-2t_2M)^{-1} B\eta_1$$
$$- 2t_1t_2\eta_1'B(D-2t_1B-2t_2M)^{-1} M\eta_1.$$

Similarly,

$$(A.4) \quad t_2\eta_2'M\eta_2 = t_2\eta_2'M(D-2t_1B-2t_2M)^{-1} D\eta_2 - 2t_2^2\eta_2'M(D-2t_1B-2t_2M)^{-1} M\eta_2$$
$$- 2t_1t_2\eta_2'B(D-2t_1B-2t_2M)^{-1} M\eta_2.$$

Thus, the expression in (A.1) becomes

$$\text{(A.5)}\quad \kappa(t_1, t_2) = \sum_{r=1}^{\infty} \frac{2^{r-1}}{r} \text{tr.}\ (t_1D^{-1}B+t_2D^{-1}M)^{-1} + t_1\eta_1'B(I-2t_1D^{-1}B-2t_2D^{-1}M)^{-1}\eta$$

$$+ t_2\eta_2'M(I-2t_1D^{-1}B-2t_2D^{-1}M)^{-1}\ \eta_2$$

$$- 2t_1t_2(\eta_1-\eta_2)'B(I-2t_1D^{-1}B-2t_2D^{-1}M)^{-1}\ D^{-1}M(\eta_1-\eta_2)\ .$$

Since $D = B + M$, it is easy to see that the matrix $BD^{-1}M$ is symmetric. In virtue of this property, we have

$$\text{(A.6)}\quad (t_1D^{-1}B + t_2D^{-1}M)^r = \sum_{i=o}^{r} \binom{r}{i}\ t_1^i\ t_2^{r-i}\ (D^{-1}B)^i\ (D^{-1}M)^{r-i}$$

and, for sufficiently small values of t_1 and t_2,

$$\text{(A.7)}\quad (I-2t_1D^{-1}B - 2t_2D^{-1}M)^{-1} = \sum_{i=o}^{\infty} \sum_{j=o}^{\infty} 2^{i+j}\ t_1^i\ t_2^j\ \binom{i+j}{i}\ (D^{-1}B)^i\ (D^{-1}M)^j.$$

Substituting (A.6) and (A.7) into (A.5) and after a little rearrangement, we find,

$$\text{(A.8)}\quad \kappa(t_1, t_2) = 1 + \sum_{r=1}^{\infty} 2^{r-1}\ t_1^r \left\{\frac{1}{r}\ \text{tr.}\ (D^{-1}B)^r + \eta_1'D(D^{-1}B)^r\ \eta_1\right\}$$

$$+ \sum_{r=1}^{\infty} 2^{r-1}\ t_2^r \left\{\frac{1}{r}\ \text{tr.}\ (D^{-1}M)^r + \eta_2'D(D^{-1}M)^r\ \eta_2\right\}$$

$$+ \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} 2^{r+s-1}\ t_1^r t_2^s\ \frac{(r+s-2)!}{r!s!} \Big\{(r+s-1)\ \text{tr.}\ D^{-1}G^{rs}$$

$$+ (r\eta_1+s\eta_2)'G^{rs}(r\eta_1+s\eta_2) - r\eta_1'G^{rs}\eta_1 - s\eta_2'G^{rs}\eta_2 \Big\}$$

where

$$G^{rs} = D(D^{-1}\ B)^r\ (D^{-1}\ M)^s\ .$$

Upon differentiating (A.8), we obtain

$$\text{(A.9)} \quad \kappa_{ro} = 2^{r-1} (r-1)! \left\{ \text{tr.} (D^{-1}B)^r + r\eta_1' D(D^{-1}B)^r \eta_1 \right\}$$

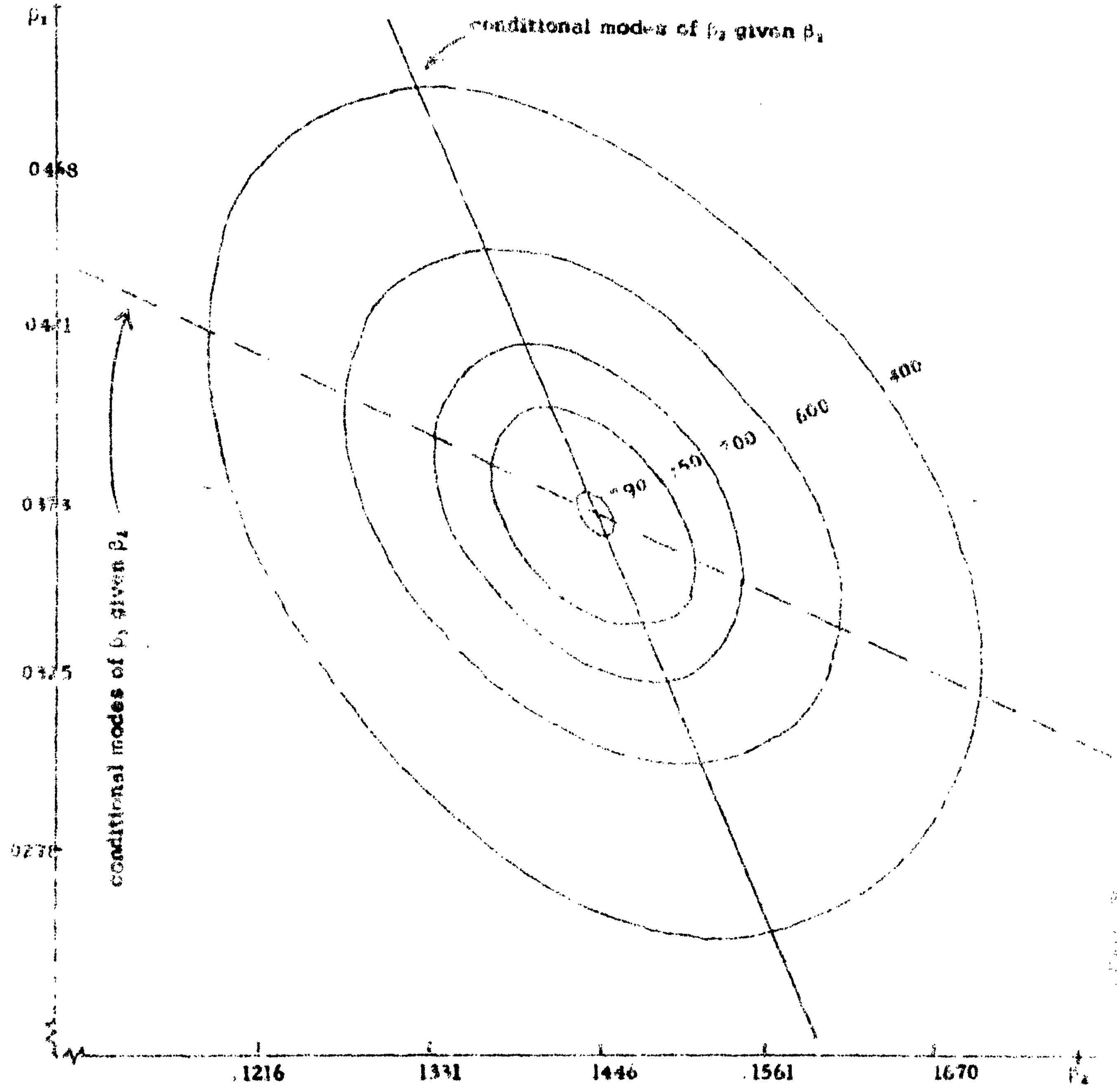
$$\text{(A.10)} \quad \kappa_{os} = 2^{s-1} (s-1)! \left\{ \text{tr.} (D^{-1}M)^s + s\eta_2' D(D^{-1}M)^s \eta_2 \right\}$$

$$\text{(A.11)} \quad \kappa_{rs} = 2^{r+s-1} (r+s-2)! \left\{ (r+s-1) \text{ tr. } D^{-1}G^{rs} + (r\eta_1+s\eta_2)'G^{rs}(r\eta_1+s\eta_2) - r\eta_1'G^{rs}\eta_1 - s\eta_2'G^{rs}\eta_2 \right\} \quad r, s \geq 1$$

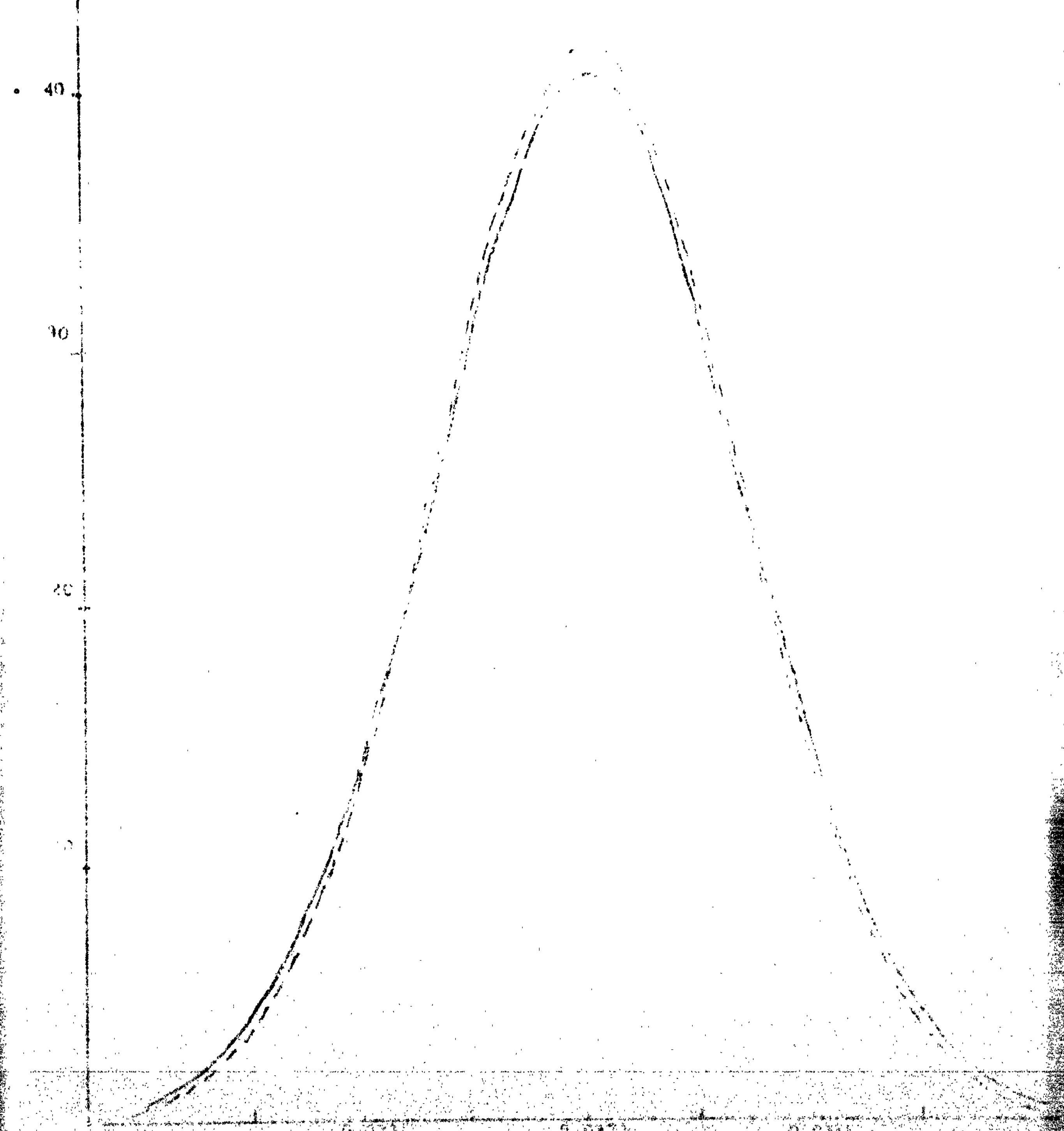
which can then be combined into the expressions given in (3.7). We note that Box (1960) has derived expressions (A.9) and (A.10) directly from the individual cumulant generating function of $Q(\beta, \hat{\beta}, B)$ and that of $Q(\beta, \tilde{\beta}, M)$, respectively.

FIGURE 1

Contours of the joint posterior distribution of β_1 and β_2



In this figure, the solid curve represents the exact distribution of [illegible] and the broken curve represents the limiting normal distribution.



References

Boot, J. C. G. and deWitt, G. M. (1960). "Investment Demand: An Empirical Contribution to the Aggregation Problem," International Economic Review, Vol. 1, 3-30.

Box, G. E. P. (1954). "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems I. Effect of Inequality of Variance in the One-Way Classification," Annals of Mathematical Statistics, Vol. 25, 290-302.

Box, G. E. P. (1960). Unpublished Lecture Notes. Department of Statistics, University of Wisconsin.

Box, G. E. P. and Tiao, G. C. (1962). "A Further Look at Robustness via Bayes Theorem," Biometrika, Vol 49, 419-433.

Box, G. E. P. and Tiao, G. C. (1963). "A Bayesian Approach to the Problem of Comparing Two Variances," Technical Report #16, Department of Statistics, University of Wisconsin.

Cook, M. B. (1951). "Bivariate κ-Statistics and Cumulants of Their Joint Sampling Distribution," Biometrika, Vol. 38, 179-195.

Cornish, E. A. (1954). "The Multivariate t-Distribution Associated With A Set of Normal Sample Deviates," Australian Journal of Physics, Vol. 7, 531-542.

Dunnett, C. W. and Sobel, M. (1954). "A Bivariate Generalization of Student's t-Distribution, With Tables for Certain Special Cases," Biometrika, Vol. 41, 153-169.

Fisher, R. A. (1961a). "Sampling The Reference Set," Sankhya, Series A, Vol. 23, 3-8.

Fisher, R. A. (1961b). "Weighted Mean of Two Samples with Unknown Variance Ratio," Sankhya, Series A, Vol. 23, 103-114.

Grunfeld, Y. (1958). "The Determinants of Corporate Investment," Unpublished Ph.D. Thesis, University of Chicago.

Jeffreys, H. (1957). Scientific Inference (2nd edition). Cambridge University Press.

Jeffreys, H. (1961). Theory of Probability (3rd edition). Clarendon Press, Oxford.

Raiffa, H. and Schlaifer, R. (1961). Applied Statistical Decision Theory, Harvard University.

Savage, L. J. (1959). "Subjective Probability and Statistical Practice," Tech. Note 59-1161, Air Force Office of Scientific Research.

Savage, L. J. (1961). The Subjective Basis of Statistical Practice, Manuscript, University of Michigan.

Savage, L. J. (1962). "Bayesian Statistics," in Decision and Information Processes, Macmillan, New York.

Theil, H. and Goldberger, A. S. (1960). "On Pure and Mixed Statistical Estimation in Economics," International Economic Review, Vol. 2, 65-78.

Theil, H. (1962). "On the Bayesian Approach to Regression Analysis," Report 6206, Econometric Institute, Netherlands School of Economics, Rotterdam.